

Listing of Claims:

1. (Previously presented) A method for retrieving information using a search engine comprising the steps of:
 - (a) retrieving a document to be indexed and temporarily storing the document in a storage device;
 - (b) determining whether relevant information is contained in the document;
 - (c) if the document contains relevant information, generating a document extract corresponding to the document by extracting a portion of the document that characterizes the document's subject content to form the document extract;
 - (d) replacing the document in the storage device with the document extract;
 - (e) decomposing the document extract into a plurality of tokens; and
 - (f) storing the plurality of tokens in a search index, wherein the search engine accesses the search index to retrieve information in one or more document extracts satisfying a search query.

2. (Previously presented) The method of claim 1, wherein the generating step (c) further comprises the steps of:

- (c1) recording positional information of the portion extracted within the document.

3. (Canceled)

4. (Previously presented) The method of claim 2, wherein the storing step (f) further

comprises:

(f1) storing the recorded positional information with the plurality of tokens.

5. (Previously presented) The method of claim 1, wherein the generating step (c) further comprises the step of:

(c1) extracting from the document a collection of sentences that are characteristic of the document's subject content to form a document summary.

6. (Previously presented) The method of claim 1, wherein the decomposing step (e) further comprises:

(e1) selecting from the document extract one of a whole sentence, a portion of a sentence, a word, and a feature.

7. (Previously presented) The method of claim 6, wherein the selecting step (e1) further comprises:

(e1i) selecting based on frequency of occurrence, word-salient-measure, proximity to the beginning of a paragraph, proximity the beginning of the document, and proximity to or position within a heading or a caption.

8. (Original) The method of claim 1, wherein the document is a web-page in the Internet.

9. (Previously presented) A computer readable medium containing programming instructions for retrieving information using a search engine comprising the instructions for:

(a) retrieving a document to be indexed and temporarily storing the document in a storage device;

(b) determining whether relevant information is contained in the document;

(c) if the document contains relevant information, generating a document extract corresponding to the document by extracting a portion of the document that characterizes the document's subject content to form the document extract;

(d) replacing the document in the storage device with the document extract;

(e) decomposing the document extract into a plurality of tokens; and

(f) storing the plurality of tokens in a search index, wherein the search engine accesses the search index to retrieve information in one or more document extracts satisfying a search query.

10. (Previously presented) The computer readable medium of claim 9, wherein the generating instruction (c) further comprises the instructions for:

(c1) recording positional information of the portion extracted within the document.

11. (Canceled)

12. (Previously presented) The computer readable medium of claim 10, wherein the storing instruction (f) further comprises the instruction for:

(f1) storing the recorded positional information with the plurality of tokens.

13. (Previously presented) The computer readable medium of claim 9, wherein the generating instruction (c) further comprises the instruction for:

(c1) extracting from the document a collection of sentences that are characteristic of the document's subject content to form a document summary.

14. (Previously presented) The computer readable medium of claim 9, wherein the decomposing instruction (e) further comprises the instruction for:

(e1) selecting from the document extract one of a whole sentence, a portion of a sentence, a word, and a feature.

15. (Previously presented) The computer readable medium of claim 14, wherein the selecting instruction (e1) further comprises the instruction for:

(e1i) selecting based on frequency of occurrence, word-salient-measure, proximity to the beginning of a paragraph, proximity the beginning of the document, and proximity to and position within a heading and a caption.

16. (Original) The computer readable medium of claim 9, wherein the document is a web-page in the Internet.

17. (Previously presented) A system for retrieving information, wherein the system includes a search engine comprising:

means for retrieving a document from a document repository and temporarily storing the document in a storage device;

an information extractor coupled to the means for retrieving, wherein the information extractor determines whether relevant information is contained in the document and if the document contains relevant information, generates a document extract corresponding to the document by extracting a portion of the document that characterizes the document's subject content to form the document extract;

means for replacing the document in the storage device with the document extract ;
a search engine indexer coupled to the storage device for decomposing the document extract into a plurality of tokens; and

a search index coupled to the search engine indexer for storing the plurality of tokens, wherein the search engine accesses the search index to retrieve information in one or more document extracts satisfying a search query.

18. (Previously presented) The system of claim 17, wherein the information extractor records positional information of the portion extracted within the document.

19. (Original) The system of claim 18, wherein the search index stores the positional information associated with the plurality of tokens.

20. (Previously presented) The system of claim 17, wherein a token of the plurality of tokens comprises one of a whole sentence, a portion of a sentence, a word, and a feature of the document.

21. (Previously presented) The system of claim 17, wherein the search engine indexer

selects the plurality of tokens based on frequency of occurrence, word-salient-measure, proximity to the beginning of a paragraph, proximity the beginning of the document, and proximity to and position within a heading and a caption.

22. (Original) The system of claim 17, wherein the document repository is the Internet and the document is a web-page.

23. (Original) The system of claim 22, wherein the means for retrieving the document is a web crawler.